Niveau(x) d'enseignement: 3 Numéro EDUQ : 8170961
Centre d'intérêt: 72

Boyer, Jean-Yves.

L'utilisation du test de closure comme instrument d'évaluation formative dans la classe de français écrit : rapport de recherche / Jean-Yves Boyer . -- Hull : Université du Québec à Hull, [1980?].
70 p. ; 29 cm.

La couv. porte en outre: PPMF. Bibliographie: p. 69-70.

Cote Dewey: 371.271/B789u Cote LC: LB/1050.46

Descripteur(s): *Français (Langue) *Langue écrite *Aire ouverte *Evaluation de l'apprentissage *Evaluation formative *Lecture *Ecriture *Test de lecture *Primaire (Niveau)

Identificateur(s) de sujet: *Test de closure *Technique close

Les étapes suivantes ont été réalisées: production d'un guide de fabrication, d'administration et de correction du test; conception d'un test de closure; réalisation d'un guide d'observation destiné à guider l'évaluation par les professeurs dont le jugement doit servir à valider le test; expérimentation de ce matériel auprès de classes du deuxième cycle du primaire. Le but recherché est de savoir si le test de closure peut servir d'instrument d'évaluation formative des habiletés en lecture et en écriture.

Douze classes de quatrième et cinquième années de cinq écoles de la région de l'Outaouais ont été impliquées dans le projet. Quatre classes ont servi à des fins de pré-expérimentation. L'essentiel du projet s'est réalisé avec la collaboration de classes à aire ouverte: deux groupes de trois classes de cinquième année (143 élèves) et un groupe de deux classes de quatrième année (52 élèves). Le choix de classes à aire ouverte permet l'évaluation des élèves par plusieurs professeurs qui les connaissent déjà suffisamment. L'analyse des résultats au test de closure a été effectuée à l'aide du programme SPSS. Le test et les guides sont reproduits en annexe.

En ce qui concerne l'élaboration, l'administration et la correction du test, les avantages que l'on reconnaît généralement à ce type d'instrument semblent confirmés et son utilisation paraît à la portée de l'ensemble des enseignants du deuxième cycle du primaire. Malgré certaines qualités docimologiques, le caractère très global du diagnostic qu'il permet lui confère toutefois une utilité limitée. On ne peut en effet assurer qu'il puisse être utilisé en vue de produire un diagnostic précis sur les diverses habiletés de lecture ou d'écriture impliquées dans l'apprentissage de l'écrit.

Type de document: d
Région(s) administrative(s): 07

Financement: Bénéficiaire - Université du Québec à Hull
Organisme(s) source(s) - MEQ - DGES
Programme(s) de subvention - PPMF
Auteur de l'analyse: ar

E D U Q 8170961

PPMF

le test de closure

Université du Québec à Hull

E D U Q 8170961

RAPPORT DE RECHERCHE

L'utilisation du test de closure comme instrument d'évaluation formative dans la classe de français écrit.

JEAN-YVES BOYER

Professeur au département des sciences de l'éducation

Université du Québec à Hull

Remerciements

La réalisation de ce projet a été rendue possible grâce à une subvention de la Direction générale de l'enseignement supérieur du ministère de l'éducation du Québec. Je tiens, de plus à remercier les enseignants des commissions scolaires Champlain, Gatineau et Outaouais-Hull qui ont collaboré à l'expérimentation du matériel produit au cours de ce projet. Ma reconnaissance s'adresse aussi à Monsieur Toussaint Fortin, responsable du PPMF, et à Madame Gisèle Chevalier, professeur au PPMF, dont l'assistance et les conseils ont constitué un précieux apport pour la réalisation de ce projet.

TABLE DES MATIÈRES

Résumé du rapport p. 3

I. **Objectifs et fondements du projet** p. 5

1. Rappel des objectifs p. 5
2. Justification de la recherche p. 5

II. **Schème expérimental** p. 8

III. **Description de la démarche** p.11

1. Choix de la population p.11
2. Élaboration du test et des directives d'administration p.12
3. Élaboration du guide d'observation p.14
4. Correction du test p.16

IV. **Résultats** p.18

1. Le test p.18
2. L'évaluation par les enseignants p.20
3. Corrélations entre le test et l'évaluation des enseignants p.21
4. Corrélations entre les habiletés et les items du test p.22

V. **Interprétation des résultats.** p.24

Appendices

A. Test de closure p.27

B. Guide d'administration du test de closure p.35

C. Guide d'observation p.39

D. Moyennes des résultats au test pour les corrections A et B p.53

E. Corrélation de chaque item avec le total A du test de closure p.55

F. Corrélations entre évaluateurs p.57

G. Corrélations entre les habiletés évaluées par les enseignants p.59

H. Corrélations entre les habiletés évaluées par les enseignants et les résultats des élèves au test de closure p.61

I. Analyse factorielle des items du test de closure p.63

J. Difficulté des items p.67

Bibliographie

p.69

Résumé du rapport

Le projet avait pour but d'adapter le test de closure aux besoins de la classe de français en mettant à la disposition des enseignants les moyens de l'utiliser pour l'évaluation de l'écrit.

La démarche suivie comprenait:

a) la production d'un guide de fabrication, d'administration et de correction du test de closure.

b) la production d'un test de closure.

c) la production d'un guide d'observation destiné à guider l'évaluation des professeurs dont le jugement devait, dans le cadre du projet, servir à valider le test de closure.

d) l'expérimentation au deuxième cycle de tout le matériel qui vient d'être énuméré et sa mise au point en fonction des résultats de l'expérimentation.

Les résultats se sont avérés positifs en ce qui concerne la possibilité d'adapter cette forme de test au deuxième cycle. Par contre, la validation du test de closure a fourni des données qui nous amènent à formuler des réserves sur l'utilisation des résultats de cet instrument de mesure dans une perspective formative.

I. Objectifs et fondements du projet

1. Rappel des objectifs

Le projet avait pour objectif général de rendre les enseignants capables d'utiliser le test de closure dans un cadre formatif en vue d'établir le profil diagnostique des élèves en classe de français écrit au primaire. De façon plus spécifique, il devait aboutir à la production d'un guide d'utilisation du test de closure pour les enseignants du deuxième cycle.

Le projet prévoyait également une expérimentation du test de closure à différents niveaux: élaboration du test, administration en classe, validation et interprétation des résultats.

2. Justification de la recherche

2.1 Aspects théoriques

Jusqu'à ce jour, la recherche sur les tests de closure s'est surtout effectuée dans les pays de langue anglaise et notamment aux Etats-Unis. Le rapport de recherche publié par G. De Landsheere (1973) présente une bibliographie qui renvoie presque exclusivement à des ouvrages et des activités en langue anglaise. De plus, la plupart des données disponibles sur le sujet portent sur l'application du test de closure à l'évaluation d'une langue seconde bien que certaines études aient établi des rapports de similitude entre les résultats obtenus en l_1 et en l_2 (Anderson 1979). Il apparaît donc qu'on manque de données concernant l'application du test de closure en langue maternelle à une population scolaire québécoise du niveau primaire.

Par ailleurs, on peut distinguer deux courants principaux dans la recherche américaine sur le test de closure. Le premier commence avec la création de

la technique par Taylor en 1953. Il a donné naissance à un grand nombre de recherches dont une bonne part était consacrée à l'analyse de la validité du test de closure comme instrument de mesure de diverses habiletés linguistiques.

Comme le mentionne Alderson (1978, 2) il se dégageait de ces études un consensus général qui reconnaissait le test de closure comme un instrument valide de la compréhension de l'écrit. Toutefois, certaines voix discordantes sont venues nuancer ce point de vue. Parmi les plus écoutées, on peut citer celle de J. Carroll (1959) et, plus récemment, de J. W. Oller (1974). Le premier suggère que le test de closure mesure la maîtrise d'habiletés proprement linguistiques plutôt que la compréhension du texte (incluant le contenu); le second met en relief une corrélation élevée qu'il a constatée entre les résultats du test de closure et ceux de la dictée.

L'autre courant de recherche, d'origine plus récente, se trouve résumé dans un article de J. C. Alderson (1979), qui, tenant compte à la fois de l'application du test de closure en l_1 et l_2, remet en question certaines de ses qualités pratiques et théoriques . Selon cet auteur, le test de closure ne mesurerait pas les habiletés de lecture au-delà des limites de la phrase. Par ailleurs, une des qualités pratiques du test de closure, comme la possibilité de produire un test par simple suppression de mots espacés régulièrement, serait illusoire. Chaque type d'espacement et de répartition des trous produirait un test différent dont la validité serait à établir à chaque fois.

2.2 Aspects pratiques

Pour l'enseignant en salle de classe, le test de closure tel que présenté par De Landsheere (1973) ou Anderson (1976), apparaît d'abord comme un outil pédagogique économique. Facile à construire, il se corrige aisément et permet d'obtenir, dans des conditions courantes, des indices de fidélité élevés (.80 ou plus).

Il reste à voir dans quelle mesure des résultats obtenus à si bon compte sont vraiment utiles pour l'évaluation formative. Si l'hypothèse la plus optimiste se vérifiait, non seulement le test de closure donnerait une mesure globale de l'habileté en lecture ou en écriture, mais on pourrait prévoir qu'une analyse des items du test permettrait de distinguer des catégories d'items qui mesurent une habileté particulière plutôt que telle autre.

II. Schème expérimental

L'expérimentation impliquait tous les niveaux que nous venons d'énumérer, mais la partie la plus complexe concernait la vérification de la validité du test de closure comme instrument d'évaluation diagnostique. Notre hypothèse de départ se formulait comme suit:

> "Le test de closure est un instrument de mesure valide pour l'évaluation des habiletés de lecture et d'écriture ainsi qu'à certaines sous-habiletés qu'elles comportent".

Cette hypothèse accordait, au départ, un maximum de valeur au test de closure. Le schème expérimental utilisé pour la vérifier prévoyait une vérification distincte pour chacun des points suivants:

1. Le test de closure fournit une mesure globale de la compréhension en lecture.

2. Le test de closure permet une mesure globale de l'expression écrite.

3. Le test de closure peut servir à mesurer l'habileté à comprendre un texte.

4. Le test de closure peut servir à mesurer l'habileté à comprendre une phrase.

5. Le test de closure peut servir à mesurer la compréhension du vocabulaire.

6. Le test de closure peut servir à mesurer l'habileté à composer un texte.

7. Le test de closure peut servir à mesurer l'habileté à comprendre une phrase.

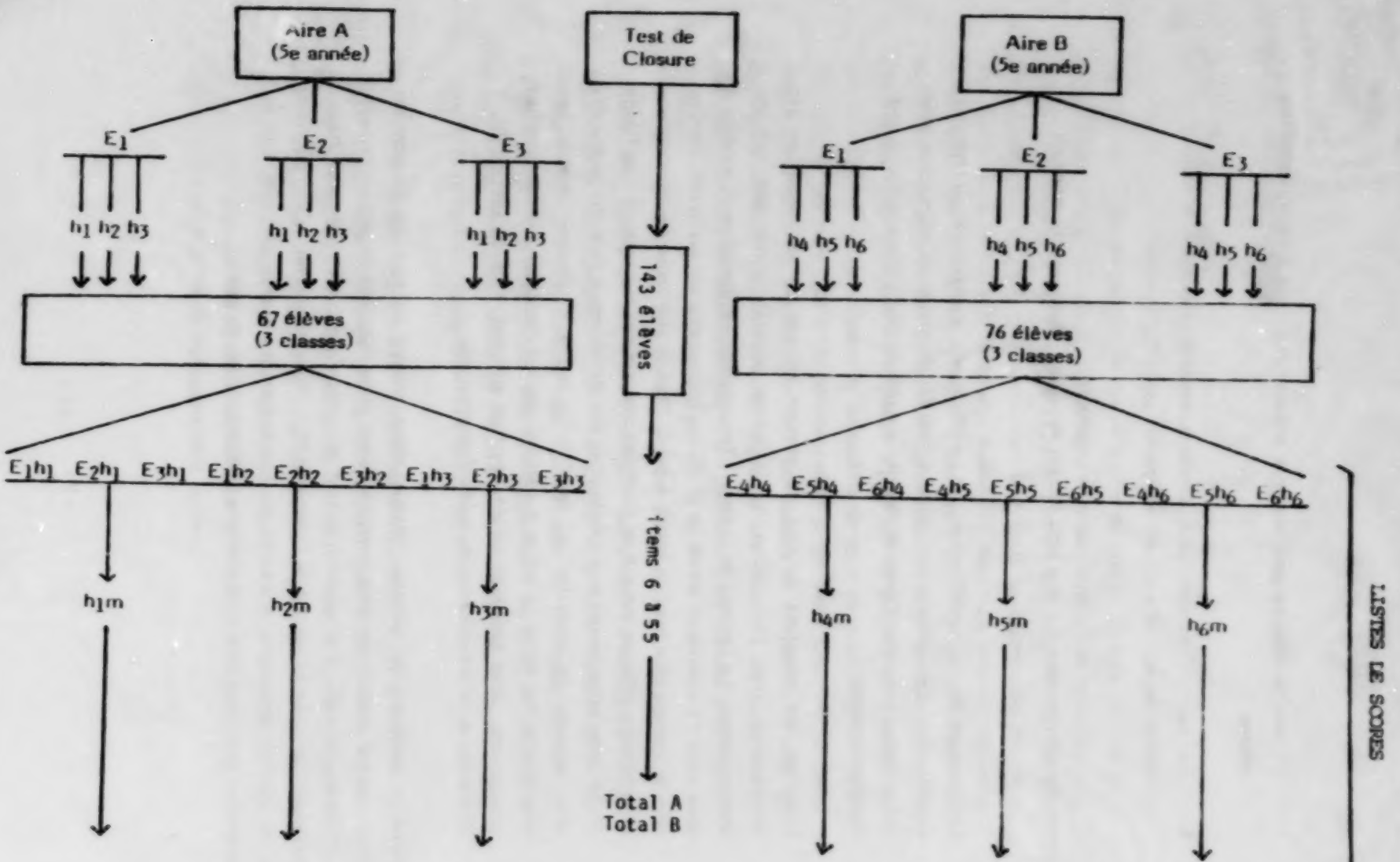
8. Le test de closure peut servir à mesurer l'habileté à utiliser le vocabulaire.

La démarche décrite dans les pages qui suivent recouvre deux aspects majeurs:

a) L'utilisation et la modification par vérifications successives du matériel expérimental (test, directives, guide d'observation, mode de correction avec des groupes) avec des élèves du deuxième cycle au cours d'une phase de pré-expérimentation.

b) L'analyse des résultats du test de closure en vue de sa validation par comparaison avec l'évaluation de la même population par des groupes d'enseignants. Le tableau de la page 10 représente l'application du schème à deux aires de cinquième année (A et B), les quatrièmes années ayant, comme nous le verrons plus loin, un statut à part. Chaque aire regroupe trois classes et le nombre d'élèves indiqué ne tient compte que des cas retenus à la fin de l'expérimentation (après soustraction des cas d'absences à l'une ou l'autre des deux séances de passation du test). La lettre E désigne chacun des enseignants des aires, la lettre h désigne une des habiletés évaluées par les enseignants. Ces habiletés seront décrites au point 3 de la section III, à l'occasion de la présentation du guide d'observation.

Comme le démontre le tableau, chaque élève soumis au test de closure est concurremment évalué par trois enseignants sur trois habiletés de lecture (aire A) ou d'écriture (aire B). Par ailleurs, le test de closure est corrigé de deux façons différentes (décrites au point 4 de la section III). Pour l'essentiel, la validation du test de closure consistera à mettre en corrélation les scores attribués par les enseignants avec ceux qu'auront produits l'administration du test.


Aire A
(5e année)
Test de
Closure
Aire B
(5e année)
E1
E2
E3
h1 h2 h3
h4 h5 h6
67 élèves
(3 classes)
143 élèves
76 élèves
(3 classes)
E1h1 E2h1 E3h1 E1h2 E2h2 E3h2 E1h3 E2h3 E3h3
E4h4 E5h4 E6h4 E4h5 E5h5 E6h5 E4h6 E5h6 E6h6
h1m
h2m
h3m
h4m
h5m
h6m
items 6 à 55
Total A
Total B
LISTES DE SCORES

III. Description de la démarche

1. Choix de la population

Au départ, nous voulions vérifier les conditions d'utilisation du test de closure au primaire. Compte tenu de la nature même du test de closure qui implique un échantillonnage au hasard des éléments d'un texte, on admet généralement qu'une cinquantaine d'items sont requis au minimum (Bloomer, 1966, cité par De Landsheere, 1973,20). En choisissant le rythme de suppression des mots de 1 sur 5, on en arrive à un texte d'au moins 250 mots. La longueur du texte créait alors une contrainte qui nous a conduit à situer l'expérimentation au niveau du deuxième cycle. En tout, douze classes de 4e et 5e années ont été impliquées dans le projet, dont quatre au niveau du guide d'administration. Les huit autres, soit six classes de cinquième et deux de quatrième (au total 195 élèves) ont participé à l'expérimentation proprement dite. A l'origine, onze classes devaient y participer, soit, six classes de cinquième et cinq de quatrième années. Des événements imprévus ont contraint certaines enseignantes d'une aire de quatrième année (comprenant trois classes) à se retirer du projet.

En tout, cinq écoles ont participé au projet. Les écoles La Sablonnière (C.S. Gatineau) et Ste-Élisabeth (C.S. Champlain) ainsi que l'école Jean de Brébeuf (C.S.O.H.) ont fourni des classes pour la pré-expérimentation. L'essentiel du projet s'est réalisé avec la collaboration des aires de cinquième et quatrième années de l'école Le Progrès (C.S. Champlain) et de l'aire des cinquièmes années de l'école Côte-du-nord (C.S.O.H.).

Le choix de classes regroupées en aire ouverte était justifié par le procédé de validation que nous avons adopté. Ce fait explique que certaines aires n'aient pas été retenues en raison de l'exigence de base: chacun des enseignants devait effectivement enseigner à tous les élèves et les connaître suffisamment pour se prononcer sur leurs habiletés à lire ou à écrire (selon les aires). En fait, dans chacune des aires impliquées, cette condition se réalisait à des degrés divers selon le mode d'enseignement retenu.

Mais, dans chaque cas également, les enseignants ont modifié dans une mesure variable leur mode de collaboration, de façon à accentuer les contacts de chaque professeur avec l'ensemble des élèves et à rendre plus facile l'observation des habiletés à évaluer.

Une aire de quatrième année a participé à l'expérimentation, mais elle ne regroupait que deux classes et deux enseignants. De plus, en cours d'expérimentation, un des enseignants a dû se retirer de l'enseignement pour des raisons de santé tandis que l'autre continuait néanmoins la tâche d'évaluation de l'aire. Pour cette aire (la seule au niveau de quatrième) nous ne disposons donc que de données incomplètes dont nous ne ferons usage qu'occasionnellement.

2. Élaboration du test et des directives d'administration

Pour l'élaboration du test et des directives d'administration, nous avons d'abord procédé à un premier ajustement du test en fonction de la réaction des élèves (en particulier, du point de vue du niveau de difficulté) et de la vérification de l'efficacité des directives du guide d'administration.

Deux tests de closure ont été utilisés. L'un était un extrait modifié d'un conte de A. Daudet, la "Chèvre de Monsieur Séguin", l'autre était construit à partir d'un texte rédigé par l'auteur du projet. Le premier test est apparu nettement trop difficile. Le second a été retenu pour modification*.

Du point de vue de l'indice de difficulté, deux critères antagonistes entraient en ligne de compte:

1. Un test de closure plus difficile (indice inférieur à .50) discrimine mieux les niveaux de performances des candidats.

2. Le test de closure difficile a un effet anxiogène qui le rend rébarbatif pour les élèves et risque de contaminer les résultats.

*Ce test, dans sa forme définitive est présenté à l'appendice A.

Le premier critère a été établi par des recherches antérieures (Oller, 1972; Alderson, 1978, 11). Le second nous est apparu nettement à l'expérience et cette constatation venait confirmer celle de Rankin (1965,252). Ce qui nous a amené à chercher un niveau de difficulté aussi rapproché que possible de .50.

Par égard au deuxième critère, nous avons cru utile de modifier les directives d'administration de façon à susciter un climat aussi détendu que possible*. Nous avons également adopté un modèle de présentation du test qui tenait compte de ce critère. D'une part, nous avons prévu de donner à l'élève les cinq premières réponses du test; ce qui a pour effet de le familiariser avec la tâche à accomplir. Par ailleurs, un texte de 250 mots auquel on ajoute les 25 mots des réponses données, représente une masse d'écrit assez impressionnante pour un élève de quatrième ou de cinquième année. Cette considération nous a amené à intervenir au niveau de la présentation du test et au niveau des modalités de passation.

Pour ce qui est de la présentation, nous avons choisi le modèle reproduit (en format réduit) à l'appendice A et qui amène l'élève à inscrire sa réponse en dehors du texte. Si nous avions invité l'élève à inscrire sa réponse dans les trous du texte, il aurait fallu prévoir des espaces assez grands pour permettre à l'élève d'écrire lisiblement, sous peine de voir apparaître des problèmes de correction. Par contre, des espaces trop larges auraient accru la longueur apparente du texte, ce qui aurait encore eu pour effet d'impressionner défavorablement l'élève.

Si on fait exception des réserves émises par Alderson (1979), l'ensemble des auteurs acceptent les conclusions de MacGinnitie (1961, 130-131) sur le fait que la distance entre les trous du test de closure ne semble pas influencer le niveau de difficulté du test à partir d'un certain seuil. Ce seuil correspondrait à la suppression d'un mot sur cinq. Toujours dans le but de réduire au minimum la longueur du test, nous avons choisi cette formule d'effacement du cinquième mot.

---

*Le guide d'administration contenant ces directives est reproduit à l'appendice B.

Pour ce qui est des critères servant à mesurer l'espacement des trous, il importait seulement de disposer d'une règle qui soit d'application simple et accessible à tout utilisateur éventuel. Cette règle peut se résumer comme suit: chaque trou est précédé de quatre mots. Le mot se définit alors de façon mécanique: tout ensemble de caractères contigus précédé et suivi d'un espace blanc dans le texte. Il y a deux exceptions à la règle: les mots élidés comptent pour un mot (l'apostrophe équivalant à un espace blanc) et les groupes incluant un trait d'union comptent pour un mot unique si le trait d'union est stable, c'est-à-dire, si on retrouve l'ensemble tel quel dans le dictionnaire.

A l'usage, l'application de ces règles n'a soulevé aucun problème particulier.

3. Élaboration du guide d'observation

3.1 Fonction et utilisation du guide d'observation

Tout le processus de validation du test de closure repose sur la comparaison des résultats du test avec l'évaluation de trois enseignants qui connaissent chacun des élèves impliqués dans l'expérimentation. Il importait donc que l'évaluation des enseignants constitue un point de repère aussi précis que possible en même temps que justifié du point de vue de son contenu.

L'alternative aurait consisté à comparer, selon la méthode classique, le test de closure à d'autres tests déjà validés. Deux raisons nous ont fait écarter ce procédé:

a) Du point de vue pratique, il aurait fallu, dans un intervalle de temps restreint, imposer aux élèves, en plus des deux séances requises par le test de closure, une série de tests portant sur chacune des habiletés évaluées. Or, nous voulions éviter ces séances répétées d'évaluations. Celles-ci, en plus de chambarder le fonctionnement normal des classes

créent un contexte artificiel, étranger au climat dans lequel doit se dérouler une évaluation formative aussi intégrée que possible au processus d'apprentissage.

b) Il nous paraissait difficile, sinon impossible, de trouver ou de produire des tests spécialisés qui mesurent exactement les habiletés en fonction desquelles nous voulions vérifier la validité du test de closure.

En fait, nous partions d'une problématique théorique sur la validité du test de closure et c'est en fonction d'elle que nous avons défini les habiletés susceptibles d'être évaluées par cet instrument. Or, l'essentiel de cette problématique situe la validité présumée du test à trois niveaux possibles: le texte, la phrase et l'unité inférieure à la phrase. Dans le dernier cas, nous avons fait correspondre l'unité au mot. Ce choix arbitraire du point de vue théorique, se justifie pour des raisons pratiques, l'enseignant étant déjà habitué par sa pratique pédagogique à distinguer ces trois niveaux d'unité linguistique.

Le guide d'observation présenté à l'appendice C représente la forme définitive d'un outil qui a été remanié à la suite de rencontres et de discussions avec les enseignants qui devaient l'utiliser. Il faut le considérer surtout comme un aide-mémoire et comme un outil de travail dont le contenu a été encore explicité et précisé en cours d'expérimentation. La période d'observation prévue et réalisée effectivement s'étendait sur un trimestre (de septembre à décembre). Ces mois d'observation devaient permettre à chaque enseignant d'acquérir une connaissance plus précise de chacun des élèves de l'aire et à chacune des équipes d'enseignants (trois par aire, en cinquième année) de se familiariser avec l'utilisation du guide et de se concentrer sur le contenu des habiletés décrites. L'auteur du projet intervenait périodiquement pour vérifier l'état de cette concertation.

3.2 L'attribution des scores

Il était entendu que les enseignants évaluaient les élèves de leur aire à partir du rendement observé au cours des deux dernières semaines d'observation qui précédaient l'administration du test. Chaque enseignant attribuait alors à chaque élève un score pour chacune des trois habiletés observées.

Le guide propose un barème de 1 à 100 avec possib.lité d'utiliser des fractions de point. A l'usage, on s'est rendu compte qu'il était peu réaliste de songer à des nuances allant au-delà du demi-point. La difficulté d'évaluer un aussi grand nombre de candidats à l'intérieur d'une même liste a été en partie résolue par la procédure progressive proposée par le guide et qui prévoit un regroupement préalable des élèves en sous-catégories de dix en commençant par les cas situés aux extrémités de la liste.

4. Correction du test

Le test de closure peut être corrigé de différentes manières selon qu'on exige le terme exact contenu dans le texte original ou qu'on accepte un équivalent. Alderson (1976, 28), résumant les résultats des recherches antérieures, conclut que les deux modes de correction donnent des résultats hautement équivalents en terme de corrélation. Anderson (1978, 15) ajoute que les habiletés mesurables par les deux procédés ne diffèrent pas.

Dans la mesure où l'on souhaite mettre à la disposition des enseignants un outil d'évaluation aussi simple et économique que possible, le mode de correction qui exige la restitution exacte du mot supprimé semble tout indiqué. Un tel procédé évite les décisions difficiles et améliore l'objectivité de la correction.

Néanmoins un problème reste à résoudre: qu'entend-on par restitution exacte du mot? De Landsheere (1973, 15) sans les justifier, proposait les règles suivantes:

"On accepte les fautes d'orthographe d'usage qui n'empêchent pas d'identifier le mot avec certitude."
"Les fautes d'orthographe grammaticale entraînent la perte du point."

Ces règles sont apparues, à l'usage, d'application facile et pratique, surtout si on a prévu, comme nous l'avons fait, un formulaire de réponse qui limite les risques d'écriture illisible. Ces règles ont aussi l'avantage de tenir compte d'un élément essentiel de la communication (qu'il s'agisse de lecture ou d'écriture): le sens. L'orthographe d'usage, moyennant la condition incluse dans la règle, n'affecte pas la signification du mot. L'orthographe grammaticale, par contre, dans la mesure où elle exprime des relations entre les mots, les propositions ou les phrases, est évidemment porteuse de signification.

Nous avons cependant prévu deux systèmes de correction en vue de vérifier l'équivalence des résultats. Chaque test a donc été corrigé selon les règles que nous venons d'énoncer pour produire un total A. Un total B a été obtenu pour les mêmes tests avec une correction qui tolérait les erreurs d'orthographe grammaticale. Dans tous les cas la bonne réponse donnait deux points et la mauvaise réponse (selon le critère A ou B) équivalait à 0. Les deux totaux (A et B) équivalaient donc directement à un pourcentage (50 réponses valant 2 points).

IV. Résultats

La correction des tests comme l'évaluation parallèle réalisée par les enseignants ont été faites manuellement. L'analyse statistique portant sur le test lui-même, sur les données fournies par les enseignants ainsi que sur les diverses corrélations et les analyses factorielles ont été obtenues par traitement informatique utilisant un programme SPSS. Le détail de ces données se retrouve dans les tableaux présentés aux appendices D à J.

1. Le test

Considérons d'abord les caractéristiques qui se dégagent de l'administration du test, à l'exclusion de sa valeur diagnostique qui sera analysée plus bas (au point 3). D'abord sa difficulté. Notre objectif était, pour les raisons énumérées plus haut, d'obtenir un résultat voisin de .50 pour l'ensemble des groupes impliqués dans l'expérimentation. Or, les résultats en pourcentage ont été les suivants*:

Ensemble des 5e années:	50.1
Aire A:	55.7
Aire B:	45.2
Aire C (4e années):	36.2

Nous pouvons tirer deux conclusions de ces premières données. La première est d'ordre pratique. Ces résultats semblent confirmer l'hypothèse que l'adaptation d'un test de closure par un enseignant à un niveau de difficulté convenant à sa classe ne soulève pas de problèmes particulier. S'il est vrai que nous avons eu recours à une série d'essais préliminaires pour ajuster la difficulté de notre test au niveau des cinquièmes années, il faut prendre en

---

* On ne tient compte ici que du total A. On trouvera à l'appendice D, le tableau des moyennes pour les corrections A et B.

considération le fait que nous ne disposons ni de l'expérience ni de la connaissance que chaque enseignant a de sa propre classe. Si un expérimentateur venant de l'extérieur peut ajuster avec une précision satisfaisante (moyennant quelques vérifications) le niveau de difficulté de son test, on peut présumer que la même tâche sera au moins aussi facile pour l'enseignant qui prépare le test pour ses propres élèves.

Une deuxième conclusion semble se dégager qui rejoint ce que De Landsheere (1973,70) constatait à la suite de sa propre expérimentation: le test de closure apparaît particulièrement sensible au niveau pédagogique, l'écart entre les moyennes des quatrièmes et cinquièmes années étant significatif (d'après le t de student) à un niveau de probabilité très élevé (.001). Quant à l'écart entre les moyennes des deux aires de 5e année, (significatif avec p=.05) il correspond à la prévision des enseignants. Les deux aires correspondaient en effet, à des milieux socio-culturels différents et, avant même que soient connus les résultats des tests, les enseignants de l'aire B prévoyaient des résultats inférieurs pour leurs propres élèves.

Le pouvoir de discrimination du test a aussi été analysé d'un autre point de vue: celui de la corrélation de chacun des items avec le total. On s'accorde habituellement à reconnaître qu'un indice situé entre .15 et .20, obtenu par cette méthode, correspond à un niveau de discrimination satisfaisant (Brown, 1976, 280). La moyenne des indices obtenus pour l'ensemble des items est de .36 (N=195). Près des trois quarts des items ont un indice égal ou supérieur à .30 et deux seulement, sur cinquante, ont un indice inférieur à .15*.

La fidélité du test a été calculée de deux façon. D'abord par la méthode du "split half" et ensuite au moyen du coefficient α. Le coefficient produit par la première méthode (avec la correction de Spearman-Brown) correspond à .80 et le second à .83. Ces coefficients sont obtenus à partir des résultats des 143 élèves compris dans les deux aires de cinquième année. Si on inclut

---

* La liste des coefficients de corrélation item/total est donnée à l'appendice E.

les résultats des quatrièmes années chacun des indices augmente de quelques points et on obtient alors .82 et .86. Le total B obtenu à partir d'une correction qui ne tenait pas compte des erreurs d'orthographe grammaticale a évidemment correspondu à des résultats plus élevés. Mais, comme prévu, on constate une corrélation élevée entre les deux séries de résultats (total A et total B), soit: .84 pour l'aire A, .94 pour l'aire B et .96 pour l'aire C. Pour l'ensemble des cinquième années (aires A et B), la corrélation est de .92 et pour les trois aires réunies (195 élèves): .94.

2. L'évaluation par les enseignants

Chacun des enseignants devait évaluer indépendamment des autres les élèves de son aire sur les trois habiletés de lecture ou d'écriture rattachées à l'aire. Il y eut donc concertation sur le mode et les objectifs de l'évaluation, mais la tâche d'observation et d'évaluation fut faite séparément de façon à produire trois évaluations indépendantes. Il en est résulté des corrélations modérées entre les correcteurs pour les différentes habiletés comme permet de le constater le tableau de l'appendice F.

Le score final accordé à chaque élève pour chacune des habiletés était une moyenne des trois scores attribués par les trois enseignants. Dans tous les cas sauf un, la corrélation entre ces scores moyens et les résultats au test de closure est supérieure aux corrélations des scores attribués par chaque enseignant avec le résultat du test.

On constate, par ailleurs, des corrélations très élevées entre les diverses habiletés évaluées, surtout en ce qui concerne l'expression écrite. Ces corrélations sont de l'ordre de .90, .84 et .86 en compréhension et de .98 et .99 en expression (voir le tableau correspondant à l'appendice G). Nous reviendrons dans la section suivante sur l'interprétation de ces chiffres.

3. Corrélations entre le test et l'évaluation des enseignants.

La comparaison entre les résultats du test de closure et de l'évaluation des enseignants s'est faite à deux niveaux que nous considérerons successivement.

3.1 Corrélations entre les habiletés et le résultat total au test.

Nous ne tiendrons compte ici que du total A, qui correspond au premier mode de correction décrit précédemment. Le tableau de l'appendice H montre qu'il n'y a pas de divergence appréciable entre les corrélations obtenues avec le total A ou le total B pour ce qui est des habiletés d'écriture et que des écarts plus sensibles se manifestent pour les habiletés de lecture.

Les corrélations élevées déjà observées entre les différentes habiletés de lecture et d'écriture permettaient de soupçonner l'existence de corrélations assez voisines entre chacune de ces habiletés et le résultat du test. C'est ce qui se produit effectivement. Voici ces corrélations:

Compréhension		
	Texte:	.58
	Phrase:	.66
	mot:	.55
Expression		
	Texte:	.77
	Phrase:	.78
	Mot:	.78

3.2 Corrélations de la compréhension et de l'expression avec le résultat total au test.

Les scores attribués à chaque élève pour la compréhension et l'expression représentent la moyenne des scores que l'élève a obtenus aux trois habiletés

correspondant à chacun de ces domaines. La compréhension correspond donc à (h1+h2+h3)/3 et l'expression à (h4+h5+h6)/3.

Il est entendu que les scores ainsi obtenus ont une signification limitée et ceci à deux points de vue.

a) La signification des termes "compréhension" et "expression" ne peut pas excéder la somme de ce que recouvrent les définitions des habiletés sous-jacentes dans le guide d'observation.

b) La moyenne obtenue suppose arbitrairement que le savoir-lire ou le savoir-écrire sont constitués à parts égales par les habiletés sous-jacentes (au niveau du texte, de la phrase et du mot). Cette supposition ne repose sur aucune certitude théorique. Nous considérerons donc les résultats obtenus comme des approximations plutôt que comme des données rigoureuses. Par ailleurs, le fait que pour les trois habiletés impliquées dans chacun des savoirs on obtienne des niveaux de corrélation très voisins, incite à croire que si l'on faisait varier la pondération de chacune dans la composition du savoir global, le résultat final n'en serait pas tellement affecté.

Une fois faites les réserves qui précèdent, on constate que le test de closure apparaît relié plus étroitement aux habiletés d'écriture plutôt que de lecture. Soit: .63 pour la lecture et .78 pour l'écriture.

4. Corrélations entre les habiletés et les items du test.

Les données qui précèdent concernaient les rapports existant entre divers niveaux d'habiletés en écrit et le total du test de closure. On peut aussi envisager la possibilité que certaines parties seulement du test soient reliées de façon significative à certaines des habiletés définies précédemment.

Nous avons donc procédé à une analyse factorielle portant sur les 50 items du test de closure. Les résultats de cette analyse devaient permettre de faire ressortir, à l'intérieur du test, des sous-ensembles d'items ayant en commun un degré de saturation élevé pour un facteur déterminé. Ces sous-ensembles auraient pu, par la suite, être mis en rapport avec chacune des habiletés de lecture et d'écriture.

Les tableaux de l'appendice I présentent les résultats de l'analyse factorielle avec trois et six facteurs. Il en ressort clairement que les items du test ne peuvent être regroupés autour d'aucun de ces facteurs avec un degré de saturation appréciable. Il est donc apparu impossible de relier un groupe d'items particulier à une des habiletés de lecture ou d'écriture.

Si les résultats de l'analyse factorielle s'étaient avérés positifs, nous avions prévu, en partant d'une analyse des erreurs, de définir en termes linguistiques les groupes d'items reliés à telle ou telle habileté particulière. Compte tenu des résultats obtenus, cette dernière analyse devenait sans objet dans la perspective du présent projet où il s'agissait de vérifier l'existence , pour l'enseignant, de moyens à la fois simples et valides d'interpréter les résultats d'un test de closure administré à ses élèves.

V. Interprétation des résultats et implications pratiques.

Des éléments d'interprétation ont été fournis au fur et à mesure que nous décrivions les étapes et les résultats de notre projet dans les sections qui précèdent. Il s'agira ici d'en faire la synthèse.

En ce qui concerne l'élaboration, l'administration et la correction du test de closure, il semble bien que les avantages que l'on reconnaît généralement à cet instrument se soient confirmés et que son utilisation soit à la portée de l'ensemble des enseignants du deuxième cycle moyennant l'utilisation d'un matériel assez simple du type de celui que nous avons décrit à la section III et que nous présentons aux apprendices A et B.

Par contre, au niveau de l'interprétation des résultats, les données sont apparues décevantes. On ne peut, en effet, assurer que le test de closure puisse être utilisé en vue de produire un diagnostic précis sur les diverses habiletés de lecture et d'écriture impliquées dans l'apprentissage de l'écrit. Même au niveau de l'habileté globale à lire, contrairement à l'opinion courante, le test ne peut pas être utilisé sans réserve puisqu'il est apparu, lors de l'expérimentation, qu'il mesure au moins autant l'habileté à écrire que le savoir-lire.

En dépit de certaines qualités docimologiques qui sont ressorties de l'expérimentation, comme la fidélité du test et sa sensibilité aux divers niveaux pédagogiques des groupes, le caractère très global du diagnostic que permet le test de closure lui confère une utilité limitée.

Il est entendu qu'une correction attentive des réponses au test peut fournir une foule d'indications sur les habiletés de chaque élève en orthographe, en grammaire et en compréhension. Mais ce genre de correction enlève au test la valeur pratique et économique qu'on lui attribuait par hypothèse au départ et équivaut à la correction successive d'un ensemble de tests qui porteraient sur l'orthographe, la compréhension, le vocabulaire et la grammaire.

A partir de ces constatations, il n'est plus évident que le test de closure soit un outil vraiment utile dans la perspective des objectifs que nous avions en vue, soit, mettre à la disposition des enseignants un instrument pratique et efficace pour l'évaluation formative. L'élaboration d'un guide d'utilisation perd donc une bonne part de sa raison d'être et nous jugeons conséquent d'en suspendre la réalisation et la diffusion dans la mesure où l'utilité du test de closure reste problématique.

Il importe toutefois de préciser la portée de ces conclusions. Elles ne valent naturellement qu'à l'intérieur des cadres de notre expérimentation et en tenant compte, en particulier, de la population impliquée.

On constate, par exemple, qu'au niveau des habiletés d'écriture, la corrélation entre les évaluations des enseignants pour chaque habileté est très élevée. On peut supposer qu'au moins une part de cette corrélation est due au fait que les enseignants ont eux-mêmes eu du mal à distinguer les différents niveaux d'évaluation. Cette supposition se justifie par le fait, qu'au témoignage des enseignants eux-mêmes, cette difficulté correspond à la situation d'enseignement. Les élèves des classes impliquées ont, en effet, peu d'occasions d'élaborer des textes de longueur appréciable. Leur apprentissage de l'écrit, était, au moment de l'évaluation, encore largement restreint au niveau de la phrase. A la lumière de cette constatation, il devient compréhensible que le test de closure n'ait pas pu distinguer entre des habiletés déjà mal distinguées au niveau de l'apprentissage. Et dans une mesure moindre, les mêmes considérations valent pour la lecture.

Il semblerait donc pertinent de nuancer nos conclusions de la manière suivante: le test de closure ne permet pas un diagnostic précis sur les habiletés définies au départ lorsqu'on l'applique à une population scolaire dont le niveau et le type d'apprentissage de l'écrit correspond à celui de notre échantillon. Ce qui laisse subsister la possibilité que le test de closure puisse donner des résultats différents dans le cadre d'un programme d'apprentissage où, par exemple, la compréhension ou l'élaboration du texte occupe une place plus importante par rapport aux objectifs traditionnels de maîtrise des éléments de la langue comme la grammaire et le vocabulaire.

Cette dernière hypothèse poura revêtir un intérêt croissant à mesure que se réalisera l'implantation du nouveau programme de français. Mais, pour l'instant et compte tenu de la situation actuelle de l'enseignement du français écrit au primaire, les conclusions de notre recherche ont surtout une valeur de mise en garde contre l'utilisation abusive de la technique de closure comme instrument d'évaluation.

APPENDICE A

Test de closure

TEST DE CLOSURE

Version expérimentale

FORME 1d*

*Reproduction en format réduit de l'original

Tu en connais des _1_ : As-tu déjà pensé _2_ tout

ce que tu _3_ faire quand tu parles? _4_ y a

d'abord _5_ mots que tu emploies. _6_ sont

nombreux! Si tu _7_ en être sûr essaie _8_ peu

de les compter. _9_ y a les noms _10_ , les noms

communs, les _11_ , les adjectifs, les pronoms, _12_ mots

invariables... Si tu _13_ les écrire tous, tu _14_ aurais

pas assez de _15_ une journée. Il y _16_ aussi toutes

les phrases _17_ tu fais. Elles sont _18_ nombreuses

encore. Songe à _19_ ce que tu dis _20_ une seule

semaine. Répètes- _21_ toujours les mêmes phrases _22_ les

mêmes mots? Bien _23_ que non! Même quand _24_ répètes

pour mieux te _25_ comprendre, il est rare _26_ tu

reprennes la même _27_ mot pour mot. Tu _28_ dis:

1. 2.

3. 4.

5. 6.

7. 8.

9. 10.

11. 12.

13. 14.

15. 16.

17. 18.

19. 20.

21. 22.

23. 24.

25. 26.

27. 28.

beaucoup de mots 29 tu peux faire un 30 nombre	29.	30.
de phrases. Mais 31 peux faire mieux encore. 32 sais,	31.	32.
en plus, parler 33 cent manières différentes. Tu 34 que	33.	34.
j'exagère? Pas 35 tout. Parles-tu 36 directeur de	35.	36.
l'école 37 tu le fais avec 38 petite soeur?	37.	38.
Essaie donc 39 peu, dans ta tête 40 parler à ta	39.	40.
petite 41 . Ensuite, répète exactement les 42 phrases	41.	42.
en pensant que 43 parles au Directeur. Alors? 44 -tu	43.	44.
convaincu maintenant?		
Mais, 45 plus de savoir parler, 46 sais lire et	45.	46.
écrire. 47 conséquent, il y a 48 -six lettres que tu	47.	48.
49 capable de reconnaître et d' 50 pour	49.	50.
former les mots 51 la langue française. Et 52 y	51.	52.
en a beaucoup.		

53 me diras peut-être que 54 fais des

fautes en 55 ces mots. Mais il y a le

dictionnaire! Et si tu connais tes lettres, tu peux

lire tout ce que tu veux dans le dictionnaire. C'est

comme si tu en savais autant que ces messieurs qui l'ont

fait ce gros livre.

53. 54.

55.

Feuille de réponse - Forme 1d

1. choses*	2. à*	29. et	30. grand
3. sais*	4. Il*	31. tu	32. Tu
5. les*	6. Ils	33. de	34. penses
7. veux	8. un	35. du	36. au
9. Il	10. propres	37. comme	38. ta
11. verbes	12. les	39. un	40. de
13. voulais	14. n'	41. soeur	42. mêmes
15. Toute	16. a	43. tu	44. es
17. que	18. plus	45. en	46. tu
19. Tout	20. en	47. Par	48. vingt
21. tu	22. et	49. es	50. écrire
23. sûr	24. tu	51. de	52. Il
25. faire	26. que	53. Tu	54. tu
27. phrase	28. sais	55. écrivant	

---

* Les réponses marquées d'un astérisque sont données aux élèves au cours de l'exercice de préparation au test proprement dit.

APPENDICE B

Guide d'administration du test de closure

1. L'enseignant distribue les textes. Il a soin de retourner les copies sur les pupitres et de demander aux élèves de ne pas les lire avant d'en recevoir la consigne.

2. Quand toutes les copies sont distribuées, il invite les élèves à lire en silence ce qui est écrit sur la page en essayant de comprendre le plus possible le contenu de l'ensemble du texte. Il leur signale qu'une petite partie seulement des mots ont été effacés, et qu'en lisant les autres, ils peuvent se faire une opinion assez précise du contenu de la page. Il leur annonce aussi qu'ils auront cinq minutes pour prendre connaissance du texte et qu'ils ne doivent rien écrire avant d'en recevoir le signal.

3. Après les cinq minutes de lecture, l'enseignant explique aux élèves la nature exacte de la tâche qu'on attend d'eux. Ils devront deviner le mot qui manque dans l'espace laissé libre. Il n'y a pas de rapport entre la dimension du trou dans le texte et la longueur du mot. Le mot manquant peut être court (comme un " l' ") ou long (comme un nom composé du type "rendez-vous"). Cependant il n'y aura jamais plus d'un mot à trouver pour chacun des trous.

4. Pour aider les élèves à mieux comprendre la tâche qui leur est demandée, l'enseignant trouve avec eux la réponse pour les cinq premiers trous. Il doit toutefois veiller à ce qu'il reste aux élèves exactement 50 trous à combler individuellement.

5. L'enseignant offre aux élèves de répondre à leurs questions sur la procédure à suivre. Il fournira toutes les explications susceptibles d'aider les élèves, hormis bien sûr, celles qui leur donneraient la réponse à un des 50 derniers items du test.

6. A la première séance, l'enseignant informe les élèves qu'ils disposeront d'une demi-heure pour répondre au test, mais qu'on ne s'attend pas à ce qu'ils aient le temps de répondre à toutes les questions. Une deuxième séance aura lieu au cours de laquelle ils auront tout le temps nécessaire pour achever le test.

7. Les élèves passent le test.

8. Après une demi-heure, l'enseignant leur demande d'inscrire leur nom à l'endroit prévu sur leur copie. Il recueille ensuite les copies en s'assurant que le nom de l'élève y figure bien.

9. A la deuxième séance, on prévoit encore quelques minutes de lecture du texte et une période consacrée aux questions. Les copies sont recueillies au fur et à mesure que les élèves ont terminé, mais pas avant que le premier quart d'heure ne soit écoulé.

APPENDICE C

Guide d'observation

1. BUT DU GUIDE

Ce guide d'observation a pour but de vous faciliter l'évaluation de la langue écrite de vos élèves. Il couvre six domaines d'apprentissage que nous avons tâché de définir d'une façon qui soit à la fois simple et précise.

Nous vous proposons d'observer chez chacun de vos élèves les manifestations de six habiletés correspondant à la distinction traditionnelle entre la compréhension et l'expression ainsi qu'à trois niveaux de maîtrise du langage: le texte, la phrase et le mot. Votre grille d'observation ressemblera donc, dans ses grandes lignes, à ce qui suit:

	Compréhension	Expression
Texte		
Phrase		
Mot		

La liste des domaines d'observation devient alors:

1. comprendre un texte
2. produire un texte
3. comprendre une phrase
4. produire une phrase
5. comprendre des mots
6. produire des mots (pour s'exprimer)

Les pages qui suivent préciseront les limites et les implications de chacun de ces six domaines. Par la suite, vous trouverez (au point 3) des directives sur la manière de classer vos élèves en fonction des critères expliqués au point 2.

2. DÉFINITION DES DOMAINES D'OBSERVATION

2.1 Comprendre un texte

Question:

Jusqu'à quel point l'élève peut-il répondre à une question clairement posée sur la signification directe de l'ensemble d'un texte?

Explication:

a) "Une question clairement posée."

La difficulté ne doit, en aucune manière, résider dans la formulation ou le contenu de la question. Si l'élève était incapable de répondre correctement, il devrait être sûr que son impuissance à fournir une réponse convenable serait due à son incompréhension du texte lui-même et non de la question.

b) "la signification directe".

La question posée doit porter sur le contenu effectivement exprimé dans le texte et non sur un sens dérivé par l'interprétation ou le point de vue subjectif du questionnaire. Par exemple, si le texte à comprendre était la fable "Le lièvre et la tortue", vous ne pourriez pas demander à l'élève s'il y a dans son entourage des gens qui ressemblent à ces deux animaux. Vous pourriez, par contre, demander laquelle des deux bêtes est la plus douée pour se déplacer rapidement. Vous pourriez même poser la question: "D'après la fable, est-ce qu'il est plus important de pouvoir aller vite ou de faire des efforts réguliers et continus?", parce que le texte lui-même propose explicitement un point de vue sur ce sujet.

c) "l'ensemble du texte".

La question doit porter sur une idée importante et non sur un point de détail. Souvent - mais pas toujours - l'idée maîtresse d'un texte est exprimée dans le

titre. Une question convenable pourrait consister à faire trouver par l'élève un titre adéquat.

Suggestions:

Tous les élèves peuvent répondre à une question du type de celles que nous venons de décrire. Il suffit que la question porte sur un texte suffisamment court et facile. Pour réaliser son classement, et distinguer entre les différents niveaux d'habileté, l'enseignant pourra tenir compte des critères suivants:

a) La longueur du texte.

b) La difficulté du texte

- quant à sa forme (vocabulaire, phrase et style)
- quant à son contenu (sujet, idées).

Toutes choses étant égales par ailleurs, un meilleur lecteur pourra comprendre le sens général d'un texte plus long et plus difficile.

2.2 Produire un texte

Question:

Jusqu'à quel point l'élève peut-il, sur un sujet qui lui est bien connu et qui l'intéresse, produire un texte significatif et bien structuré?

Explication:

a) "un texte significatif".

C'est-à-dire un texte qui a du sens, dont les idées sont reliées entre elles et avec le sujet proposé à l'élève ou choisi par lui.

b) "bien structuré".

Cette qualité a trait, pour l'essentiel, au plan du texte. Le texte a un début et une fin. Entre les deux, le déroulement est cohérent (qu'il soit logique, chronologique ou autre). Ce qui implique que le texte doit avoir une longueur minimale, même si la longueur n'est pas un critère en soi. On pourrait fixer à sept ou huit lignes cette longueur minimale.

Remarque: On évite ici de mesurer certaines qualités, importantes par ailleurs, comme la créativité, l'originalité, l'inspiration. Pour des raisons de méthode, on se limite strictement, dans le cas présent, à l'habileté qui consiste à construire un texte bien structuré et cohérent.

2.3 Comprendre une phrase

Question:

Jusqu'à quel point l'élève peut-il répondre à une question clairement formulée sur la signification d'une phrase, dans la mesure où cette signification est reliée à la structure syntaxique?

Explication:

a) "Une question clairement formulée".

cf. l'explication donnée en 2.1.

b) "signification (...) reliée à la structure syntaxique".

En utilisant les mêmes mots, on peut construire plusieurs phrases dont les sens sont différents ou même contraires. En ajoutant quelques mots-outils (prépositions, conjonctions) on veut varier presque à l'infini la signification d'une phrase. Il est question ici de ce sens des phrases qui tient aux relations entre

les mots ou les groupes de mots (propositions). Cette relation peut être exprimée soit par l'ordre des mots, soit par les mots grammaticaux qui les relient. On évaluera donc l'habileté de l'élève à comprendre le sens d'une phrase plus ou moins complexe, sans que n'intervienne (dans la mesure du possible) la compréhension du vocabulaire.

Suggestions:

a) La longueur de la phrase donne déjà un bon indice de sa complexité. Néanmoins, il y a (pour des phrases de longueur égale) des différences de difficulté reliées à la nature des relations entre les mots et les groupes de mots.

Exemples:

Une phrase passive, pourra être plus difficile à comprendre que sa forme active. Une phrase qui cumule les formes interrogative, négative et passive sera encore plus difficile.

Une phrase qui comporte une subordonnée concessive sera plus difficile qu'une autre qui contient seulement une relative.

Une phrase contenant plusieurs subordonnées sera généralement plus difficile qu'une autre qui n'en a qu'une, etc.

b) Pour éviter que la compréhension du vocabulaire ne vienne compliquer l'évaluation de la compréhension de la phrase, vous pouvez observer l'habileté de l'élève à comprendre des phrases dans lesquelles vous êtes sûrs que le vocabulaire est tout à fait familier.

c) Pour éviter que la compréhension du contexte n'interfère dans la compréhension de la phrase, vous pouvez, dans la mesure du possible, ne tenir compte que de la compréhension des phrases isolées. Si un contexte est nécessaire, vous l'exposez vous-même oralement, avec toutes les précisions nécessaires.

2.4 Produire une phrase

Question:

Jusqu'à quel point l'élève, quand il écrit, utilise des phrases bien construites?

Explication:

"des phrases bien construites".

a) Il s'agit de correction syntaxique. Cette correction se situe au niveau des relations entre les parties de la phrase. C'est-à-dire qu'il n'y a pas d'erreurs d'accord entre les mots; pas d'erreurs de régime (bonne préposition entre le verbe et son complément); pas d'erreurs dans les rapports entre les diverses propositions, notamment entre la principale et la ou les subordonnée(s).

b) Il s'agit aussi de variété dans la construction. Un élève pourrait produire des phrases comportant peu de fautes, mais elles seraient toutes construites sur un modèle unique du type sujet - verbe - complément. Un autre élève qui aurait le même nombre d'erreurs mais utiliserait des constructions plus variées et plus complexes, serait jugé supérieur.

Remarque: L'orthographe est impliquée dans la correction syntaxique, dans la mesure où elle concerne l'accord entre les mots (orthographe grammaticale). On ne tiendra pas compte ici de l'orthographe d'usage.

2.5 Comprendre des mots

Question:

Jusqu'à quel point l'élève peut-il, par des synonymes, des périphrases ou quelqu'autre moyen, montrer qu'il comprend les sens précis d'un mot employé hors contexte?

Explication:

"le sens d'un mot employé hors contexte".

Nous excluons ici l'habileté à découvrir (partiellement ou complètement) le sens d'un mot en utilisant le contexte où il se trouve.

Les critères pertinents sont

a) l'étendu du vocabulaire (nombre de mots connus)

b) la précision du vocabulaire (habileté à rendre le sens exact d'un mot ou à distinguer clairement les différents sens qu'il peut recouvrir).

2.6 Produire des mots (pour s'exprimer)

Question:

Jusqu'à quel point l'élève, lorsqu'il écrit, utilise-t-il des termes propres et variés?

Explication:

a) "des termes propres".
Il s'agit ici de la précision et de la convenance des mots dans le contexte où ils sont utilisés.

b) "des termes (...) variés".

On évalue la variété du vocabulaire, en mesurant, dans un texte la proportion de termes différents par rapport au nombre total de mots employés. En d'autres termes, plus les mêmes mots reviennent souvent, moins le vocabulaire est varié.

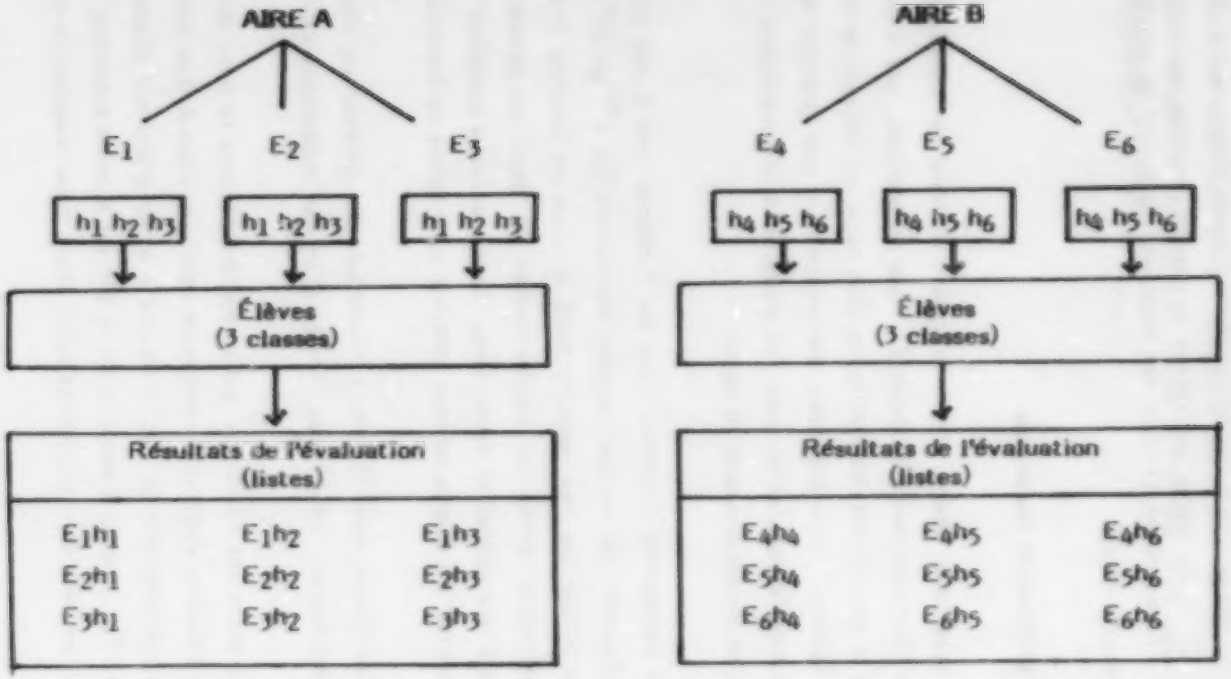
3. ATTRIBUTION DES SCORES

La démarche que nous vous proposons se veut aussi simple que possible, compte tenu du nombre d'élèves à évaluer et de la nature des habiletés à mesurer.

3.1 Définition des tâches

Au terme des trois mois d'observation prévus par le projet, chaque enseignant devra attribuer à chacun des élèves de son aire, un score qui correspond à sa performance à chacune des habiletés observées. Pour l'attribution de ce score, l'enseignant tiendra compte autant que possible du niveau de performance auquel l'élève est parvenu au cours des deux dernières semaines de la période d'observation dans l'ensemble des activités de lecture ou d'écriture à l'occasion desquelles cette performance a pu se manifester. Il importe, en effet que l'évaluation des enseignants précède de peu la passation du test de closure par les élèves.

Il faut signaler que les habiletés sont réparties entre les différentes aires, de façon à alléger la tâche des évaluateurs. Ainsi, chaque enseignant aura à attribuer un score pour trois habiletés et non six. Le schéma de la page suivante représente la répartition des tâches entre les aires. Les symboles E_1, E_2, E_3,E_4, E_5, E_6,y représentent respectivement chacun des enseignants de l'aire. Les symboles h_1, h_2, h_3, h_4, h_5, h_6 correspondent à chacune des habiletés prévues au point 2 du guide d'observation. Les formules E_1h_1, E_1h_2, etc. désignent les listes de scores produites par chacun des enseignants pour chacune des habiletés et pour chacun des élèves.


AIRE A
E1
E2
E3
h1 h2 h3
h1 h2 h3
h1 h2 h3
Élèves
(3 classes)
Résultats de l'évaluation
(listes)
E1h1
E1h2
E1h3
E2h1
E2h2
E2h3
E3h1
E3h2
E3h3
AIRE B
E4
E5
E6
h4 h5 h6
h4 h5 h6
h4 h5 h6
Élèves
(3 classes)
Résultats de l'évaluation
(listes)
E4h4
E4h5
E4h6
E5h4
E5h5
E5h6
E6h4
E6h5
E6h6

3.2 Le barème

Vous attribuez vos scores à l'intérieur d'une échelle allant de 1 à 100. Le meilleur élève aura 100. Vous notez les autres en tenant compte de leurs performances respectives. Vous pouvez utiliser des fractions de points et attribuer des scores égaux à différents élèves.

3.3 Procédure d'attribution des scores

Nous vous suggérons, étant donné le grand nombre de candidats à noter, de procéder par regroupements successifs. Vous déterminez, par exemple, ceux qui entrent dans le groupe des dix meilleurs, puis ceux qui doivent se retrouver parmi les dix plus faibles. Vous procédez successivement par groupes de dix, jusqu'au moment où toute la population aura été intégrée à une catégorie (la dernière ne comprenant pas nécessairement 10 noms).

Une fois vos catégories formées, vous les reprenez une à une pour classer les élèves qui y figurent les uns par rapport aux autres (du 1^{er} au 10^{e}). Autant que possible vous évitez les "ex aequo"; mais si vous ne trouvez vraiment aucune raison de distinguer entre les performances de deux ou plusieurs candidats, résignez-vous à leur donner le même rang. Souvenez-vous toutefois qu'il vous sera possible d'exprimer les écarts les plus minimes sous forme de fractions de points.

Quand tous les élèves sont classés à l'intérieur des catégories, vous procédez à l'attribution des scores. Vous tenez compte alors des indications suivantes:

1) Les écarts entre les candidats qui se succèdent dans la liste doivent, dans la mesure du possible, avoir une cohérence relative; c'est-à-dire que si le 27e est nettement meilleur que le 26e, il pourra y avoir entre eux plusieurs points de différence. Par contre, si selon vous, le 26e est à peine supérieur au 25e, il n'y aura entre ces deux derniers élèves qu'un écart d'une fraction de point.

2) Si le premier de la liste a nécessairement 100, le dernier, lui, n'a pas forcément un. Il est cependant souhaitable que vous utilisiez la plus grande portion possible de l'échelle.

3) Pour ajuster votre répartition des scores aux contraintes de l'échelle, vous devrez sans doute passer par une période de tâtonnement, surtout en ce qui concerne le respect de la cohérence des écarts. Nous vous proposons alors de travailler avec un fichier. Chacune des fiches contiendrait alors, au terme de votre démarche, le nom d'un élève, le numéro de sa catégorie, son rang dans la catégorie et son score.

APPENDICE D

Moyennes des résultats au test pour les corrections A et B.

	Corr. A		Corr. B	
	M	σ	M	σ
Aire A (N=67)	55.73	10.91	73.19	11.93
Aire B (N=76)	45.18	13.74	62.29	17.22
Aire C (N=52)	36.15	14.28	49.19	17.19
Aires A et B (N=143)	50.13	13.52	67.40	15.89

APPENDICE E

Corrélation de chaque item avec le total A du test de closure (N=195)

ITEM	CORR.	ITEM	CORR.
6	.37	31	.45
7	.24	32.	.33
8	.46	33	.41
9	.15*	34	.20
10	.36	35	.51
11	.43	36	.38
12	.26	37	.45
13	.09**	38	.44
14	.16*	39	.48
15	.36	40	.41
16	.18*	41	.33
17	.47	42	.30
18	.22	43	.41
19	.51	44	.38
20	.42	45	.50
21	.35	46	.54
22	.40	47	.14*
23	.39	48	.51
24	.43	49	.19*
25	.41	50	.38
26	.36	51	.43
27	.50	52	.27
28	.24	53	.37
29	.19*	54	.26
30	.39	55	.45

*P= .001 dans tous les cas sauf ceux marqués d'un astérisque pour lesquels P est égal ou inférieur à .05. Le double astérisque désigne l'unique cas qui fait exception avec P=.09.

APPENDICE F

Corrélations entre évaluateurs*

	Lecture (Aire A : N=67)	Écriture (Aire B : N=76)
TEXTE	E_1/E_2 : .65 E_1/E_3 : .52 E_2/E_3 : .70	E_4/E_5 : .52 E_4/E_6 : .61 E_5/E_6 : .48
PHRASE	E_1/E_2 : .45 E_1/E_3 : .65 E_2/E_3 : .45	E_4/E_5 : .55 E_4/E_6 : .63 E_5/E_6 : .50
MOT	E_1/E_2 : .45 E_1/E_4 : .64 E_2/E_3 : .40	E_4/E_5 : .51 E_4/E_5 : .52 E_5/E_6 : .51

* Pour tous les cas figurant dans le tableau, P = .001.

APPENDICE G

Corrélations entre les habiletés évaluées par les enseignants

Lecture	Texte / Phrase : .90
(N=67	Texte / Mot : .84
P=.001)	Phrase / Mot : .86

Écriture	Texte / Phrase : .99
(N=76	Texte / Mot : .98
P=.001)	Phrase / Mot : .98

APPENDICE H

Corrélations entre les habiletés évaluées par les enseignants et les résultats des élèves au test de closure
(Total A et Total B)

	LECTURE		ÉCRITURE	
	Total A	Total B	Total A	Total B
TEXTE	.58	.54	.77	.78
PHRASE	.66	.60	.78	.79
MOT	.55	.52	.78	.79

APPENDICE I

Analyse factorielle des items du test de closure (3 et 6 facteurs).

Items	Facteur 1	Facteur 2	Facteur 3
6	.16373	.66017	-.28415
7	.21964	.09194	-.01599
8	.42947	.09819	.25669
9	.09153	.66382	-.13962
10	.21161	.24580	.07535
11	.29870	.34991	.05156
12	.03307	.11999	.21374
13	.02958	-.01650	.18972
14	.01901	.09815	.03585
15	.07623	.16850	.34338
16	.04290	.13312	.01038
17	.29436	.12865	.18372
18	.15854	.12903	-.02310
19	.23139	.20240	.47697
20	.12402	.36931	.27452
21	.11046	.10206	.23832
22	.01868	-.03791	.45568
23	.26830	.16340	.28758
24	.45661	.08546	.10493
25	.20171	.20495	.24970
26	.36910	.09645	.16010
27	.26414	.40465	.14463
28	.15972	.17017	.07283
29	-.05196	.16590	.05819
30	.10536	.14450	.30670
31	.38871	-.03746	.29248
32	-.16817	.58244	.15524
33	.35157	.07240	.14445
34	.22327	-.00143	.08901
35	.39175	.25573	-.03209
36	.14435	.13945	.19557
37	.05350	.35927	.49235
38	.37384	.00886	.09905
39	.59363	.07458	.22261
40	.32193	.31305	.07314
41	.40747	-.07479	.21217
42	.17856	.14330	.21970
43	.33582	-.07680	.20193
44	34519	.20201	.00891
45	.21003	.35069	.13440
46	.47886	.12361	-.04437
47	.14611	.02732	.00680
48	.23141	.08628	.17743
49	.25020	.06526	-.20185
50	.12581	.39719	.08953
51	.33243	.14331	.09011
52	-.04808	-.15647	.30803
53	-.09486	.66120	.15930
54	.26851	-.06973	-.09413
55	.21200	.22423	.14399

Items	Facteur 1	Facteur 2	Facteur 3	Facteur 4	Facteur 5	Facteur 6
6	.16466	.63408	-.10119	-.07227	.21517	-.17267
7	.10834	.08233	.03278	.10850	.17813	-.03220
8	.53871	-.02089	.29302	.07625	-.01525	-.03219
9	.18158	.67150	-.06293	.00971	-.00262	-.14889
10	.23292	.20406	.05964	.01547	.11619	.08436
11	.35328	.32049	.06626	.11366	.01593	-.10162
12	-.09316	.10873	.18666	.00232	.25401	.35105
13	.00725	-.04541	.25951	.02910	-.02652	.01211
14	.11412	.01069	.22956	-.23296	.00297	-06823
15	.23444	.13418	.10098	.04297	-.10019	.34210
16	.17467	.12590	-.00402	.02602	-.18913	-.11881
17	.29545	.07264	.10413	.06634	.14882	.18150
18	.16686	.09662	.16788	.04794	-.04978	-.28973
19	.24665	.12851	.46698	.12166	.01699	.18188
20	.21324	.41106	.01175	.25877	-.12885	.22508
21	.04654	.00292	.61246	-.08508	.06761	-.10985
22	.14939	-.07070	.08446	.07241	-.10402	.61113
23	.24633	.09441	.33429	.09130	.08128	.08885
24	.41479	.01785	.18685	.17964	.09830	-.14557
25	.17247	.21332	.07655	.23197	.04830	.21940
26	.47648	.02571	.03331	.08908	.01740	.09512
27	.32704	.33107	.16317	-.02541	13690	.10092
28	.13590	.14807	.10813	.06457	07127	-01137
29	-.03455	.16843	.02691	-.02015	.01027	.12722
30	.03933	.09321	.39563	.04415	.09984	.13624
31	.28809	-.02914	.07286	.42841	.03713	.16593
32	-.01077	.56069	.15499	-.11431	-.06870	.11361
33	.26954	.03370	.12496	.15713	.17022	.06389
34	.09096	-.02540	.09775	.08063	.24098	.11292
35	.41292	.16217	.04597	-.06677	.25708	-.00552
36	.15473	.11833	.02921	.03942	.13650	.31095
37	.09068	.31828	.52929	.12343	-.10057	.14170
38	.15029	.05131	.00469	.45870	.17250	-.01811
39	.72339	-.07565	.19739	.09633	.08516	-.00132
40	.40336	.25702	.04988	.05198	.06455	-.00555
41	.22098	-.04656	.10621	.58794	.01651	-.09001
42	.22246	.16924	-.00195	.28033	-.11296	.11833
43	-.04827	.00321	.08764	.75685	.18297	.01984
44	.17964	.15937	.03482	.02610	.50037	.16458
45	.11287	.31480	.32011	.09450	.15550	-.07427
46	.20488	.10930	.04716	.24739	.43106	-.04806
47	.12255	.02000	-.05051	.04864	.09674	.06232
48	.15040	.04335	.22130	.09881	.14284	.06263
49	-.00627	.04045	-.01813	-.02489	.54137	-.06243
50	.03643	.38269	.19845	.05003	.18833	.04764
51	.34661	.09008	.00021	.07280	.15397	.13533
52	-.07194	-.12926	.14845	.19005	-.12181	.16631
53	.00807	.70276	.07592	.07478	-.09335	.11654
54	.09884	-.10915	.04914	.01783	.29730	-.05548
55	.23644	.13932	.28557	-.06213	.09757	-.00006

APPENDICE J

Difficulté des items

Items	Aires A et B (5e années) N=143	Aire C (4e années) N=52	Items	Aires A et B (5e années) N=143	Aire C (4e années) N=52
6	.44	.46	31	.88	.64
7	.22	.12	32	.20	.10
8	.76	.50	33	.61	.23
9	.51	.15	34	.04	.02
10	.36	.21	35	.83	.52
11	.29	.12	36	.76	.52
12	.74	.62	37	.18	0
13	.01	0	38	.78	.46
14	.35	.21	39	.88	.83
15	.29	.19	40	.60	.52
16	.97	.94	41	.90	.79
17	.94	.75	42	.20	.14
18	.10	.06	43	.94	.85
19	.37	.23	44	.29	.12
20	.28	.10	45	.50	.29
21	.37	.14	46	.80	.50
22	.76	.48	47	.04	.02
23	.67	.62	48	.70	.42
24	.74	.60	49	.25	.25
25	.62	.36	50	.16	.06
26	.95	.90	51	.73	.54
27	.39	.21	52	.83	.52
28	.14	.08	53	.17	.14
29	.20	.14	54	.59	.46
30	.34	.12	55	.38	.21

BIBLIOGRAPHIE *

Alderson, J. Charles, The cloze procedure and proficiency in English as a foreign language, dans Tesol Quarterly, vol. 13, no. 2, Juin 1979, pp. 219-227.

Alderson, J. Charles, The effect of certain methodological variables on cloze test performance and its implication for the use of the cloze procedure in EFL testing, Communication présentée au 5e congrès international de linguistique appliquée, Montréal, août 1978.

Alderson, J. C., A study of the cloze procedure with native and non-native speakers of English. Thèse de doctorat non publiée; cité par J. C. Alderson 1979, pp. 226.

Anderson, Jonathan, Psycholinguistic experiment in foreign testing, St. Lucia, Queensland, University of Queensland Press, 1976.

Brown, Frederick G., Principles of Educational and Psychological Testing, Holt, Rinehart and Winston, 1976.

Carroll, J. B., A.S. Carton et C.P. Wilds, An investigation of cloze items in the measurement and achievement in foreign languages, cambridge, Mass. Harvard University, 1959.

De Landsheere, Gilbert, Le test de closure, Paris, Nathan, 1973.

---

* Cette bibliographie ne contient que les ouvrages cités dans le rapport.

MacGinnitie, W., Contextual Constraint in English Prose Paragraphes, dans Journal of Psychology, vol. 51, 1961, pp. 121-130.

Oller, J. W. Jnr, Scoring methods and difficulty levels for cloze tests of proficiency in English as a second language, dans The Modern Language Journal, vol. 56, no. 3, 1972, pp. 151-158.

Oller, J. W. Jnr, P. Atai et P. Irvine, Cloze dictation and the test of English as a foreign language, dans Language Learning, vol. 24, no. 2, 1974, pp. 245-252.

Rankin, E.F., Cloze procedure. A survey of research, dans E.L. Thurston et L.E. Hafner, The philosophical and sociological bases of reading, Fourteenth Yearbook of the National Reading Conference, Milwaukee, Wisconsin, The National Reading Conference, 1965, pp. 133-150.

Stump, Thomas A., Cloze and dictation tasks as predictors of intelligence and achievement scores, dans John W. Oller Jr. et Kyle Perkins, Language in Education, testing the tests, Rowley, Mass., Newbury, 1978.

Taylor, W.L., Cloze procedure: a new tool for measuring readibility, dans Journalism Quaterly, vol. 30, 1953, pp. 415-433.